

For the reasons set forth below, the Court will grant in part and deny in part Defendant's *Daubert* Motion. Specifically, the Court will preclude DeSouza from testifying that the cartridge

cases that were recovered matched, or came from, the firearm that was recovered, but the Court will otherwise deny Defendant's motion. Thus, the Court will permit DeSouza's testimony to include: (1) the theory of firearms and toolmark analysis; (2) the procedures he undertook to examine the cartridge cases, including inspecting and testing the firearm recovered, and his use of a comparison microscope to compare the cartridge cases to those test-fired from the recovered firearm; (3) whether the class characteristics matched; and (4) whether he found individual toolmarkings on the recovered cartridge cases to be consistent with those test-fired from the recovered weapon. Further, DeSouza may use comparison photographs to describe or show what he concludes were consistent toolmarks from the recovered cartridge cases and those that were test fired.

## **I. INTRODUCTION**

In the early morning hours of November 13, 2019, Officer Darrell Walcott ("Officer Walcott") and his partner were in their marked police vehicle facing Northside Road near Oscar Refrigeration and Furniture Center in the La Grande Princess area. *United States v. Felix*, No. 20-CR-0002, 2021 WL 3272205, at \*1 (D.V.I. July 29, 2021).<sup>1</sup> The officers were present in the area as part of the Virgin Islands Crime Initiative because, in months prior, the 911 emergency call center had been receiving numerous calls regarding gunshots in the area close to where they were parked. *Id.*

Around 3:00 a.m., Officer Walcott received reports of shots being fired as a crowd dispersed from the Starlight Nightclub, and he heard multiple gunshots. He also heard and saw a

---

<sup>1</sup> The Court takes the background factual discussion in this section from the Court's prior opinion on Defendant's Motion to Suppress. The Court provides this information solely for the purposes of this pretrial motion, ever mindful that Defendant is presumed innocent until proven guilty. Most of the facts discussed herein are alleged but not conceded or proven beyond a reasonable doubt at this stage.

vehicle traveling at a high rate of speed in the direction where he and his partner were located. Officer Walcott observed that the car driving towards him was a dark grey or silver Ford Focus with its windows down. Officer Walcott proceeded to turn on his emergency lights and siren and followed the vehicle, but the vehicle did not stop. Around this same time, another officer in a different vehicle, Officer Michael Jules (“Officer Jules”), also heard gunshots and began driving to assist Officer Walcott.

As Officer Walcott pursued, the Ford Focus arrived at an intersection and ran over a stop sign and through a fence, onto private property. Officer Walcott observed a male wearing a black baseball cap and holding a firearm exit the vehicle while it was still in motion. The individual ran around the vehicle towards the house on the property, and the car continued to move and eventually crashed into a tree. Officer Walcott lost sight of the individual. Officer Jules and other police officers then arrived at the scene, created a perimeter, and searched for the individual. Soon after, Officer Jules saw an individual on an adjacent property, who the officers arrested. Officer Walcott identified Defendant as the individual who police arrested and who he saw exiting the vehicle.

Officers later searched the area in which Defendant was apprehended and discovered a firearm and magazine. Officer Walcott also searched the car that had crashed into the tree, where he discovered two live rounds, and its vicinity, where he discovered a black firearm magazine with live rounds inside. Finally, Officer Walcott searched the area on Northside Road where he had heard shots being fired, where he found approximately eighteen 5.7 spent shell casings.

In an Indictment filed on January 21, 2020, the Government charged Defendant with three counts: “Felon in Possession of Firearm,” in violation of 18 U.S.C. §§ 922(g)(1) and 924(a)(2) (Count 1); “Felon in Possession of Ammunition,” in violation of 18 U.S.C. §§ 922(g)(1) and 924(a)(2) (Count 2); and “Possession of Firearm in School Zone,” in violation of 18 U.S.C. §

922(q)(2)(A) and 924(a)(1)(B) (Count 3).

On August 25, 2020, Defendant filed the instant *Daubert* Motion seeking to limit the testimony of DeSouza, the Government’s proposed expert witness in the field of firearms examination and comparison. (Dkt. No. 24). Defendant seeks to preclude DeSouza from testifying that the cartridge cases Officer Walcott found matched the firearm found near “the vehicle Mr. Felix is alleged to have exited.” (Dkt. No. 24 at 2). Specifically, Defendant requests that the Court: “(1) exclude testimony that the cartridge cases were fired from the particular firearm at issue; (2) limit the firearms examiner to a discussion of class characteristics; [and] (3) . . . exclude any comparison photographs that purport to show a match.” *Id.* at 5. Defendant defines “class characteristics” as “distinctive features shared by many items of the same or similar type—such as the width of a groove cut into the barrel of a firearm, or the shape of a firing pin—and are determined before manufacturing.” *Id.* (citing National Research Council, *Strengthening Forensic Science in the United States: A Path Forward*, at 152 (National Academies Press 2009)) (“2009 NRC Report”).

Defendant argues that application of Federal Rule of Evidence 702’s (“Rule 702”) “scientific and reliability requirements warrants the preclusion of all ‘identification’ or ‘match’ testimony from trial.” *Id.* at 2. Defendant asserts that such testimony is inadmissible under Rule 702 because “the methodology of firearms toolmark pattern matching is not based on reliable principles and methods and thus, lacks foundational validity.” *Id.* at 3.

## **II. BACKGROUND**

### **A. The Theory Behind Firearms Analysis**

This Court has previously considered a *Daubert* motion challenging firearms and toolmark identification. *See United States v. Wrensford*, No. 2013-CR-0003, 2014 WL 3715036 (D.V.I. July

28, 2014). In *Wrensford*, the Court incorporated a “summary of the theory underlying the discipline,” from Judge Stanley R. Chester’s learned discussion of the subject in *United States v.*

*Otero*, 849 F.Supp.2d 425 (D.N.J. 2012). The Court will reproduce that summary here:

The Third Circuit has observed that the general category of forensic identification evidence serves to connect a crime scene object or mark to the one and only source of that object or mark. Forensic toolmark identification is a discipline that is concerned with the matching of a toolmark to the specific tool that made it. Firearm identification is a specialized area of toolmark identification dealing with firearms, which involve a specific category of tools. Toolmarks are generated when a hard object (tool) comes into contact with a relatively softer object. Toolmarks associated with a firearm may occur in the commission of a crime when the internal parts of a firearm make contact with the brass and lead [or other materials] that comprise ammunition. The manufacture and use of firearms produces an extensive set of specialized toolmarks.

Toolmark identification is based on the theory that tools used in the manufacture of a firearm leave distinct marks on various firearm components, such as the barrel, breech face or firing pin. The theory further posits that the marks are individualized to a particular firearm through changes the tool undergoes each time it cuts and scrapes metal to create an item in the production of the weapon. Toolmark identification thus rests on the premise that any two manufactured products, even those produced consecutively off the same production line, will bear microscopically different marks. With regard to firearms, these toolmarks are transferred to the surface of a bullet or shell casing in the process of firearm discharge. Depending on the tool and the type of impact it makes on the bullet or casing, these surface marks consist of either contour scratch lines, known as striations (or striae), or impressions.<sup>2</sup> For example, rifling (spiraled indentations) inside of a gun barrel will leave raised and depressed striae, known as lands and grooves, on the bullet as it is fired from the weapon, whereas the striking of the firing pin against the base of the cartridge, which initiates discharge of the ammunition, will leave an impression but not striae.

Comparing a test bullet or cartridge fired from a firearm of known origin to another bullet or cartridge of unknown origin, the examiner seeks to determine congruence in the pattern of marks left on the examined specimens. This process is known as “pattern matching.” . . . An examiner observes three types of characteristics on spent bullets or cartridges: class, subclass and individual. Class characteristics are gross features common to most if not all bullets and cartridge cases fired from a

---

<sup>2</sup> “A spent bullet usually has striated marks, created as it moves through the barrel of the gun. On the other hand, a spent cartridge case can have both impressed and striated marks.” *Wrensford*, 2014 WL 3715036, at \*2 n.3.

type of firearm, for example, the caliber and the number of lands and grooves on a bullet. Individual characteristics are microscopic markings produced in the manufacturing process by the random imperfections of tool surfaces (the constantly changing tool as described above) and by use of and/or damage to the gun post-manufacture. According to the theory of toolmark identification espoused by the Association of Firearms and Toolmark Examiners (“AFTE”), individual characteristics are unique to that tool and distinguish it from all other tools. Subclass characteristics generally fill the gap between the class and individual characteristics categories. They are produced incidental to manufacture but apply only to a subset of the firearms produced, for example, as may occur when a batch of barrels is formed by the same irregular tool.

*Otero*, 849 F.Supp.2d at 427-28 (internal quotations and citations omitted).

## **B. *Daubert* Hearing Testimony**

At the *Daubert* hearing, the Court heard testimony from two witnesses: Dr. James E. Hamby (“Dr. Hamby”), “a forensic scientist specializing in firearms and toolmark identification as well as laboratory management” (Hr’g. Tr. at 25); and Dr. Nicholas Scurich (“Dr. Scurich”) who has a PhD in psychology with a specialty in quantitative psychology and is an associate professor at the University of California at Irvine, *id.* at 335-36.

### **1. Dr. James E. Hamby**

Dr. Hamby testified that he has been a firearms examiner for over fifty years and is currently self-employed as a consultant. *Id.* at 26. As a consultant, he primarily “defend[s] the science” of firearms and toolmark identification “at various legal tribunals around the country.” *Id.* Dr. Hamby testified that he has worked as a firearms and toolmark examiner (“firearms examiner”) for various entities, including the United States Army, the Illinois State Police Laboratory System, and the Indianapolis-Marion County Laboratory, where he served as director for approximately nineteen years. *Id.* at 27-28. Dr. Hamby testified that he has a PhD in Forensic Science from the University of Strathclyde in Glasgow, Scotland, and that he is a member of the “Association of Firearm and Toolmark Examiners, . . . the American Academy of Forensic

Science, the Canadian Forensic Science Society, the Midwest Association of Forensic Science, [and] the British Forensic Science Society,” among other organizations. *Id.* at 28-29. He has taught or lectured at various universities around the world and has testified in court approximately 500 times. *Id.* at 30, 32. Dr. Hamby testified that he has also trained individuals to be firearms examiners, and that he has co-authored the “AFTE Training Manual,” in addition to other publications and studies. *Id.* at 31-33. Based on his background, training, and experience, the Court admitted Dr. Hamby to testify at the *Daubert* hearing “as an expert in the area of firearms and toolmark examination.” *Id.* at 34.

At the *Daubert* hearing, Dr. Hamby discussed the theory underlying the firearm and toolmark examination field, stating that when metals are manufactured, “it’s impossible to make them identical.” *Id.* at 38. Dr. Hamby also explained the different types of marks made on fired bullets, including class, subclass, and individual marks, and provided exhibits with examples. *Id.* at 48-62.

Dr. Hamby described the process by which firearms examiners analyze bullets and cartridges, explaining that first an examiner ensures that a firearm is operable. *Id.* at 64-65. Then, the components are examined to determine if they are “the same caliber as the firearm,” because if they are not, they were not fired by that firearm. *Id.* at 65. Next, a firearms examiner will do “test-fires” and microscopic examination. *Id.* “AFTE has a requirement of two [test-fires] minimum,” but some labs will do more. *Id.* at 67. Then, the examiner will look “for replication of striae and/or impressed marks, depending on whether you have fired bullets or fired cartridge cases or both.” *Id.* at 82. An examiner will go through each piece of evidence for class characteristics and individual marks. *Id.* at 83. Dr. Hamby testified that based on a thorough examination, an examiner will make a determination, which is “thoroughly documented and recorded.” *Id.* at 66.

He also testified that firearms examinations are subject to peer review insofar as “all identification and inconclusives and eliminations [are] to be looked at by a second examiner.” *Id.* at 95-96.

Dr. Hamby testified that a firearms examiner can make certain conclusions, explaining:

[T]here are essentially four. There’s [(1)] identification or individualization, there’s [(2)] inconclusive, and there’s some gradations of that, although to me an inconclusive is an inconclusive, period. There’s [(3)] elimination or [(4)] unsuitable, because if a bullet has been fired and it smacks into the wall of the courthouse, which happens to be -- if it’s concrete and it hits and flattens out, there’s nothing left to examine.

*Id.*

For an identification, Dr. Hamby testified that there is “both objective and subjective analysis,” with the objective component being the class characteristics. *Id.* at 67. Dr. Hamby testified that there is training and experience and “you develop a skill set.” *Id.* at 84. He explained that you are “looking at the width and the depth . . . to see if you’re getting the same striated marks going across or the impressed marks . . . and its repeatable.” *Id.* at 84-85. For an identification, Dr. Hamby testified that “you see the individual characteristics and you can say, yes, I see sufficient individual characteristics to say this was identification.” *Id.* at 83. Exhibits showing identifications were provided in the form of photographs from a comparison microscope. *Id.* at 69-70; Gov’t Ex. 9.

Dr. Hamby testified that an inconclusive determination occurs when “there’s not enough identifiable material represented on the sample.” *Id.* at 70-71. He also stated that an inconclusive was not an error, in his experience, and it was “not an incorrect answer because it’s certainly not an exclusion because the class characteristics are the same. But it’s certainly not an identification either, or its not unsuitable simply because it didn’t get mangled beyond belief.” *Id.* at 76-77. Elaborating, Dr. Hamby testified an inconclusive could occur for a variety of reasons, such as a “lack of suitable marks, limited sample size, [or] damage to the components.” *Id.* at 499.



Dr. Hamby testified that an elimination occurs when the item being analyzed is different, for example “you can’t fire a .45 caliber bullet out of a .38 caliber revolver.” *Id.* at 77. Based on a Government exhibit, Dr. Hamby also testified that elimination for two cartridge cases could occur based on the shape of a firearm’s firing pin. *Id.* at 79; *see* Gov’t Ex. 13 (providing side-by-side pictures of the firing pin impressions on two separate cartridge cases).

Dr. Hamby testified that an unsuitable determination occurs when “[t]here’s nothing there. There’s nothing of value for examination.” *Id.* at 81. For example, when a bullet “smacks into a wall, a concrete wall” and “shatters into a million pieces, there’s nothing to examine.” *Id.*

Dr. Hamby testified that firearms examination is testable, despite there being some subjectivity, and it is tested frequently by virtue of research projects. *Id.* at 87. Dr. Hamby testified about a number of studies, including one he conducted. *Id.* at 87-95. Further, studies are published in peer-reviewed journals. *Id.* at 97. Based on the studies Dr. Hamby reviewed, he testified the error rates were “from zero up to one and a half maybe.” *Id.* at 99.

Specifically, Dr. Hamby testified regarding the “Ames<sup>3</sup> II,” “Validation Study of the Accuracy, Repeatability, and Reproducibility of Firearm Comparisons,” which was a “double-blind black box study,”<sup>4</sup> where the answers were unknown to researchers. *Id.* at 152; Stanley J.

---

<sup>3</sup> Ames is the name of the U.S. Department of Energy Laboratory at which the study was conducted.

<sup>4</sup> “Because the procedures for feature identification, the matching rule, and frequency determinations about features are not objectively specified, the overall procedure must be treated as a kind of ‘black box’ inside the examiner’s head.” President’s Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts: Ensuring Validity of Feature-Comparison Methods* 62 (2016), available at [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_forensic\\_science\\_report\\_final.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf). “Since the black box in the examiner’s head cannot be examined directly for its foundational basis in science, the foundational validity of subjective methods can be established *only* through empirical studies of examiner’s performance to determine whether they can provide accurate answers; such studies are referred to as ‘black-box’ studies []. In black-box studies, examiners are presented with many independent comparison problems—typically, involving ‘questioned’ samples and one or more ‘known’

Bajic et al., *Report: Validation Study of the Accuracy, Repeatability, and Reproducibility of Firearm Comparisons* (Ames Laboratory-US DOE 2020) (Tech. Rep. #IS-5207) (“Ames II”). Dr. Hamby testified that the error rate for the Ames II study was “under one, one and a half percent max.” *Id.* at 300. Dr. Hamby further testified that the error rate for “Ames I,” “A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons” was 1 or 1.1 percent. (Hr’g Tr. at 283); David P. Baldwin et al., *A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons* (2014) (“Ames I”), available at <https://www.ojp.gov/pdffiles1/nij/249874.pdf>. In his own study, Dr. Hamby test-fired 10 consecutively rifled Ruger pistols twice; provided the twenty bullets in an envelope, and then directed firearms examiners to review those against fifteen unknown bullets, all of which were also fired from one of the ten Ruger pistols. *Id.* at 161; Gov’t Ex. 22 at 6. Dr. Hamby testified he had seven inconclusives and no errors. *Id.* at 162. According to the study, this was from 697 completed tests. Gov’t Ex. 22 at 6.

Dr. Hamby also testified about five other studies. First, in the “Isolated Pairs Research Study” (“Keisler Study”), examiners were asked to review twenty pairs of cartridge cases and determine whether the cases came from the same firearm (an identification) or from a different one (an exclusion). Mark A. Keisler et al., *Isolated Pairs Research Study*, 50 AFTE J. 56, 56 (2018). For the Keisler study, Dr. Hamby testified that the reported error rate was zero, and the study was peer-reviewed. (Hr’g Tr. at 505). The study also reported 203 inconclusive responses out of 1,008 “true exclusions possible” for different source comparisons. Keisler, *supra* at 57.

Second, in “An Empirical Study/Validation Test Pertaining to the Reproducibility of

---

samples—and asked to declare whether the questioned samples came from the same source as one of the known samples. The researchers then determine how often examiners reach erroneous conclusions.” *Id.* (emphasis in original).

Toolmarks on 20,000 Bullets Fired Through M240 Machine Gun Barrels” (“Mikko Study”) examiners reviewed 164 bullets, and “all 164 were correctly ascribed to the correct answer.” Don Mikko, *An Empirical Study/Validation Test Pertaining to the Reproducibility of Toolmarks on 20,000 Bullets Fired Through M240 Machine Gun Barrels*, 45 AFTE J. 290, 291 (2013); (Hr’g Tr. at 507).

Third, in “An Empirical Study to Evaluate the Repeatability and Uniqueness of Striations/Impressions Imparted on Consecutively Manufactured Glock EBIS [(“Enhanced Bullet Identification System”)] Gun Barrels” (“Fadul Study”), “participants examined and compared the 15 questioned fired bullets to the 10 pairs of known test fired bullets . . . and determined which barrel fired the 15 questioned fired bullets.” Thomas G. Fadul, Jr., *An Empirical Study to Evaluate the Repeatability and Uniqueness of Striations/Impressions Imparted on Consecutively Manufactured Glock EBIS Gun Barrels*, 43 AFTE J. 37, 41 (2011). The EBIS barrel was specifically designed to ensure each Glock bullet fired could be traced to the pistol that fired it. *Id.* at 37. The reported error rate for the Fadul Study was 0.04 percent. *Id.* at 41.

Fourth, in the “Knife Identification Project” (“Thompson Study”), ten consecutively made knives were used to create twenty known and twenty unknown knife cuts, which were provided to examiners. Evan Thompson & Rick Wyant, *Knife Identification Project (KIP)*, 35 AFTE J. 366, 367-68 (2003). The Thompson Study states that the goal of the study “was to associate numbered, known knife blade test marks with corresponding unknown letter blocks.” *Id.* at 368. The error rate reported for the Thompson study was 0.776 percent, and “[i]nconclusive results were not considered incorrect for examiners who used this option.” *Id.* at 369.

Finally, in “The Identification of Consecutively Rifled Gun Barrels” (“Brundage Study”), examiners were each given fifteen unknown bullets and twenty known bullets, with two known

bullets coming from one of ten consecutively manufactured barrels. David J. Brundage, *The Identification of Consecutively Rifled Gun Barrels*, 30 AFTE J. 438, 438-40 (1998). According to the study, “[t]here were no incorrect answers for any of the results collected. Each examiner made the proper associations between each gun barrel and all” unknown bullets from the 30 test sets. *Id.* at 440. “Inconclusive responses were not considered incorrect” because many factors “can affect identifiability,” and there was one bullet that could not be identified. *Id.* at 440-41. Dr. Hamby also testified that he believed that the one inconclusive was from some of the markings being obliterated. *Id.*

## **2. Dr. Nicholas Scurich**

Dr. Scurich testified that he received his PhD in psychology, and he was in the subdiscipline of quantitative psychology, in which psychologists study “advanced methods to come up with tests or measurements.” (Hr’g Tr. at 335-36). Dr. Scurich explained that “the essence of [his] doctoral program was . . . how to study humans making judgments and decisions.” *Id.* at 340. He testified that he was the vice chair of the Department of Psychological Sciences at University of California, Irvine, where he also held an appointment in the “Department of Criminology, Law & Society.” *Id.* at 336. Dr. Scurich testified that he teaches undergraduate and graduate courses, one of which is in advanced research methods for doctoral students and focuses on “teaching students basic principles of research methods, [and] how to design a study that can appropriately address the question they’re interested in.” *Id.* at 337. Dr. Scurich has received awards for his work and grants from the Department of Homeland Security, and he has published articles in peer-reviewed scientific journals. *Id.* at 337, 341.

Dr. Scurich explained that studies about firearms and toolmark error rates “involve human subjects, firearm examiners, looking at objects and making judgments and decisions,” which is

within his area of expertise. *Id.* at 345. For purposes of the *Daubert* hearing, the Court accepted Dr. Scurich as an expert “in the areas of quantitative psychology, research methodology, and study design[,] to include statistics and human decision-making.” *Id.* at 349.

In his testimony, Dr. Scurich explained the different types of study designs used in firearms and toolmark studies. He explained that a “set-to-set” study design involves an examiner receiving “a known set of known bullets or cartridge cases and a set of unknown bullets or cartridge cases,” and the examiner is asked “to match up the unknown bullets or cartridge cases to the known.” *Id.* at 350. He explained that a “[s]ample-to-sample design is when the participant is given two bullets at once, essentially, and they’re asked to make a judgment: identification, elimination, or [] inconclusive.” *Id.* After the judgment is made, the bullets are put away, and an examiner is given two other bullets and asked to make another judgment. *Id.*

Dr. Scurich also testified that an “error rate is a percentage of how many of the responses are incorrect.” *Id.* A “false positive is when you’re looking at a different source comparison,” and “the bullets [or cartridge cases] did not come from the same gun” but the examiner concludes they did. *Id.* The examiner concludes there was an identification, which is false. *Id.* Whereas, a “false negative occurs when you’re looking at same source bullets,” but the examiner concludes the bullets did not come from the same gun, *i.e.*, determines it to be an elimination, which is false. *Id.* at 350-51. Dr. Scurich explained that a false positive error rate is the number of instances in which an examiner incorrectly determines a bullet or cartridge case to be a match. *Id.* at 351. Dr. Scurich also testified that for different source comparisons, where a bullet did not come from the same gun, “the correct answer is elimination.” *Id.*

Regarding study design, Dr. Scurich testified that sample-to-sample design is “superior” to set-to-set design. *Id.* at 353. Dr. Scurich explained that set-to-set studies are susceptible to various

“methodological issues.” *Id.* at 354. For example, Dr. Scurich testified that set-to-set studies do not track “the number of comparisons that were made” by an examiner, *e.g.*, how many times an examiner compared an unknown bullet to a known bullet. *Id.* at 353-54. Dr. Scurich explained that without knowing the “true number of comparisons, you can’t actually calculate a false positive error rate or a false negative error rate,” because to do so “you have to look at all the different source comparisons . . . to see how many times they got those comparisons wrong.” *Id.* at 354. Dr. Scurich also explained that set-to-set design is “problematic” because “it has a tendency to underestimate error rates in part because you can use the design to gain inferences about what the correct answer is.” Dr. Scurich further testified that a set-to-set study could be “closed” or “open.” *Id.* at 357. “[A] closed set-to-set study would be where all of the unknowns match up to at least one of the knowns.” *Id.* While “[a]n open set-to-set study would be where some of the unknowns do not match up to the knowns.” *Id.* For closed-set studies, Dr. Scurich explained that “the initial responses can be used to help fill in the subsequent responses” because “there’s an interdependency between the different comparisons, which participants could use as a sort of deductive process to help arrive at a conclusion.” *Id.* at 552.

Dr. Scurich also testified that Dr. Hamby’s study, which was set-to-set, did not: (1) calculate “repeatability,” *i.e.*, “the extent to which the same examiner looks at a set of bullets at time one and makes a judgment and reexamines the same bullet some time later . . . and reaches the same conclusion; or (2) calculate “reproducibility,” *i.e.*, when two different examiners “look at the same set of bullets and reach the same conclusion.” *Id.* at 357-59. Dr. Scurich testified that a false positive error rate could not be calculated from Dr. Hamby’s study. *Id.* at 359.

Regarding the AMES I study, Dr. Scurich testified it was a “sample-to-sample study design,” and it had not been published in a peer-reviewed journal. *Id.* at 360-61. For the AMES I

study, Dr. Scurich testified that the false positive rate was 1.01 percent. *Id.* at 362. However, Dr. Scurich was critical of the study because it counted inconclusives as correct answers. *Id.* at 363. According to Dr. Scurich, there were 2,180 different source comparisons—*i.e.*, comparisons of cartridge cases from different firearms—and 22 were labeled as identification, 735 were labeled inconclusive, and 1,421 were labeled as eliminations. *Id.* at 364. Dr. Scurich also testified that the study states that the number of samples labeled as inconclusive could not be “attributed to a large fraction of poorly marked knowns or question[ed] samples in this group.” *Id.* at 364-65. Further, Dr. Scurich explained that 45 out of the 218 examiners in the study responded “inconclusive” to every different source comparison they were given. *Id.* at 365. Overall, Dr. Scurich testified that AMES I had “22 false positives, four false negatives.” *Id.* at 367. But, he stated that if you included the inconclusives with the false positives, “the error rate would be 34.7 percent.” *Id.* at 368.

Dr. Scurich explained that when it came to inconclusives, his view was “there is a correct response, either identification or elimination, respectively, and any other response is an error because it doesn’t match up with ground truth. And so that’s why one would pool these errors together.” *Id.* Regarding inconclusives, Dr. Scurich testified that inconclusives of the kind described by Dr. Hamby exist in casework, where “evidence can be mangled or degraded or lacking any sort of information or detail such that it is appropriate and correct to call inconclusive evidence.” *Id.* at 369. Dr. Scurich explained that if it is “predetermined that there’s not enough for an identification and the person who’s taking that test says inconclusive, that’s a correct response, in my view.” *Id.* at 408. However, the AMES I study was set up “so that the bullets either came from the same gun or not.” *Id.* at 369.

Dr. Scurich testified about the AMES II study, which reported an overall false positive rate of 0.656 percent for bullets and 0.933 percent for cartridge cases. *Id.* at 375. However, Dr. Scurich

testified that AMES II also included inconclusives as correct answers, and if they were excluded, as done in the PCAST Report, the error rate would be two percent for bullets and 1.86 percent for cartridge cases. *Id.* at 377. However, Dr. Scurich testified that excluding the inconclusives is also problematic, because he asserts that “you’re effectively allowing the test subjects to pick and choose which questions they want to respond to.” *Id.* at 378. Dr. Scurich also testified that if inconclusives were included as errors, the error rates increase significantly, ranging from 23 percent to 54 percent for bullets, and 12.9 percent to 37.9 percent for cartridge cases, depending on which categories of inconclusives are included as errors. *Id.* at 379.

Dr. Scurich was also critical of the five studies discussed by Dr. Hamby. Regarding the Thompson Study on knives, Dr. Scurich explained that there is “an empirical question that could and should be tested, that is, to what extent do examiners who do very well on examining knife marks on plastic also do well on examining marks on bullets or cartridge cases.” *Id.* at 557. Dr. Scurich also explained that the Thompson Study was a closed, set-based study, for which no false positive error rate could be calculated, and it did not state the number of inconclusive responses or treat them as incorrect. *Id.* at 560-61. Rather, Dr. Scurich testified that the Thompson Study treated inconclusives as correct. *Id.* at 562. For the Fadul Study with the EBIS barrels, Dr. Scurich testified that it was also a closed, set-based study, for which no false positive error rate could be calculated. *Id.*

For the Mikko Study involving machine gun barrels, Dr. Scurich testified that a false positive error rate could not be calculated because the study “simply reports the number of answers and not the number of comparisons that led to those answers.” *Id.* at 564. Dr. Scurich further stated that with only four participants, and who they were and “what they were told to do” unknown, too much information was missing to determine whether the study results “could be generalized or



extrapolated to different populations.” *Id.* at 565. For the Brundage study on consecutively manufactured barrels, Dr. Scurich testified that it was also a closed, set-based study and no false positive error rate could be calculated. *Id.* at 565. Finally, for the Keisler Study, he testified that it was a sample-based study, and that there were four inconclusive responses when the correct answer was identification and 203 inconclusives when the correct answer was elimination. *Id.* at 566. He also explained that the Keisler Study stated that other participants were able to correctly make the four identifications that were reported as inconclusives, “therefore . . . the level of difficulty of the sample kits was not an issue.” *Id.* Dr. Scurich testified that if the inconclusives for different source comparisons were called errors, the error rate would be “about 20 percent.” *Id.* at 568. Dr. Scurich testified that, with the exception of the Keisler Study, the studies upon which Dr. Hamby relied were “not properly designed for their purpose.” *Id.* at 623.

### III. APPLICABLE LEGAL PRINCIPLES

“Under the Federal Rules of Evidence, a trial judge acts as a ‘gatekeeper’ to ensure that ‘any and all expert testimony or evidence is not only relevant, but also reliable.’” *Pineda v. Ford Motor Co.*, 520 F.3d 237, 243 (3d Cir. 2008) (quoting *Kannankeril v. Terminix Int’l, Inc.*, 128 F.3d 802, 806 (3d Cir. 1997)). “The Rules of Evidence embody a strong preference for admitting any evidence that may assist the trier of fact.” *Id.* (citation omitted). Specifically, “Rule 702, which governs the admissibility of expert testimony, has a liberal policy of admissibility. *Kannankeril* 128 F.3d at 806.

Rule 702 contains three major requirements: “(1) the proffered witness must be an expert; (2) the expert must testify about matters requiring scientific, technical or specialized knowledge; and (3) the expert's testimony must assist the trier of fact.” *Id.* (citing *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 741-42 (3d Cir. 1994) (“*Paoli II*”). The Third Circuit has labeled these

requirements as: “qualification, reliability, and fit.” *Schneider ex rel. Est. of Schneider v. Fried*, 320 F.3d 396, 404 (3d Cir. 2003) (citing *Paoli II*, 35 F.3d at 741-43).

Defendant’s *Daubert* Motion challenges the Government’s proposed expert testimony on the second prong, reliability. The Third Circuit has “concluded that ‘an expert’s testimony is admissible so long as the process or technique the expert used in formulating the opinion is reliable.’” *Kannankeril*, 128 F.3d at 806 (citing *Paoli II*, 35 F.3d at 742). An expert’s testimony “must be based on the ‘methods and procedures of science’ rather than on ‘subjective belief or unsupported speculation’; the expert must have ‘good grounds’ for his or her belief.” *Paoli II*, 35 F.3d at 742 (quoting *Daubert*, 509 U.S. at 590). “In sum, *Daubert* holds that an inquiry into the reliability of scientific evidence under Rule 702 requires a determination as to its scientific validity.” *Id.*

“While a litigant has to make more than a *prima facie* showing that his expert’s methodology is reliable . . . ‘[t]he evidentiary requirement of reliability is lower than the merits standard of correctness.’” *Pineda*, 520 F.3d at 247 (quoting *Paoli II*, 35 F.3d at 744). The Third Circuit has highlighted at least eight factors that a court can consider in assessing whether a proffered expert’s methodology is reliable:

- (1) whether a method consists of a testable hypothesis; (2) whether the method has been subjected to peer review; (3) the known or potential rate of error; (4) the existence and maintenance of standards controlling the technique’s operation; (5) whether the method is generally accepted; (6) the relationship of the technique to methods which have been established to be reliable; (7) the qualifications of the expert witness testifying based on the methodology; and (8) the non-judicial uses to which the method has been put.

*Id.* at 247-48. However, these factors are “neither exhaustive nor applicable in every case.” *Kannakeril*, 128 F.3d at 806-07. District courts have “broad discretion in determining the

admissibility of evidence, and ‘considerable leeway’ in determining the reliability of particular expert testimony under *Daubert*.” *Simmons v. Ford Motor Co.*, 132 F. App’x 950, 952 (3d Cir. 2005) (quoting *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 152-53 (1999)).

The Third Circuit has clarified that “if a court finds that an expert has employed a methodology only slightly different from a methodology that the court thinks is clearly reliable, the court should be more likely to accept the altered methodology than if it was evaluating that methodology as an original matter.” *Paoli II*, 35 F.3d at 745 n.14. “A judge should only exclude evidence if the flaw is large enough that the expert lacks ‘good grounds for his or her conclusions.’” *Wrensford*, 2014 WL 3715036, at \*9 (quoting *Paoli II*, 35 F.3d at 746). Finally, “the proponent of the evidence does not have to demonstrate that the assessments of the expert are correct—they only have to demonstrate by a preponderance of the evidence that their opinions are reliable.” *Id.* “The analysis of the conclusions themselves is for the trier of fact when the expert is subjected to cross-examination.” *Id.* (quoting *Oddi v. Ford Motor Co.*, 234 F.3d 136, 146 (3d Cir. 2000)) (internal quotation marks omitted).

#### IV. DISCUSSION

Defendant seeks to limit, but not exclude, the Government’s proposed expert witness, DeSouza, because his testimony allegedly does not satisfy the reliability prong of *Daubert*. Applying the factors set forth in *Daubert*, the Court finds that the process DeSouza used to formulate his opinion is reliable but will admit his testimony with limitations based on the Court’s concerns set forth below.

## A. Reliability Factors<sup>5</sup>

### 1. Testable Hypothesis

The “first *Daubert* reliability factor asks whether a theory or technique can be tested,” which the Supreme Court has stated “is a ‘key question’ in determining whether expert testimony should be admitted.” *United States v. Romero-Lobato*, 379 F. Supp. 3d 1111, 1118 (D. Nev. 2019) (quoting *Daubert*, 509 U.S. at 592-94). Defendant argues that the conclusions that firearms examiners draw from comparison analysis of bullets or cartridge casings are subjective and thus are not testable. In determining whether bullets or cartridge casings are a match, firearms examiners apply the standard of “significant agreement,” which is defined as when the agreement:

exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool. The statement that ‘sufficient agreement’ exists between two toolmarks means that the agreement is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility.

*Otero*, 849 F. Supp. at 431. Defendant argues that this standard permits examiners “to draw his or her own subjective conclusions about what qualifies as ‘sufficient agreement’ between test-fired samples and found ammunition, or between sets of found ammunition.” (Dkt. No. 24 at 21).

To support his argument, Defendant cites to the published studies examining and criticizing the field of firearms examination. Defendant cites to the 2009 NRC Report, which found that “the decision of the toolmark examiner remains a subjective decision based on unarticulated standards and no statistical foundation for estimation of error rates.” *National Research Council, supra* at 153-54. The 2009 NRC Report further states that “[b]ecause not enough is known about the

---

<sup>5</sup> The parties do not address the relationship of the technique to methods which have been established to be reliable, nor the non-judicial uses to which the method has been put. Accordingly, the Court has no reason to view these factors as applicable and will not address them. *See Pineda*, 520 F.3d at 247-48 (observing the factors are “neither exhaustive nor applicable in every case”).

variabilities among individual tools and guns, we are not able to specify how many points of similarity are necessary for a given level of confidence in the result.” *Id.* at 154. Defendant similarly cites the PCAST Report, which states that the “sufficient agreement” standard is circular, and “is defined as the examiner being convinced that the items are extremely unlikely to have a different origin.” PCAST Report, *supra* at 104.

Dr. Hamby’s testimony confirmed that a firearm examiner’s conclusion involved subjectivity, in that “the firearm examiner has to use her own subjective knowledge to make this judgment when conducting an examination.” (Hr’g Tr. at 622). However, Dr. Hamby testified that firearms examinations are testable and are frequently tested as demonstrated by the research studies, discussed above. *Id.* at 87-94.

Previously, in its *Wrensford* opinion, this Court found that “the theory of firearms identification consists of a testable hypothesis, notwithstanding the inherent subjectivity involved in the approach.” *Wrensford*, 2014 U.S. Dist. LEXIS 102446, at \*42. There, as here, the Court based its decision on expert testimony regarding validation studies which concluded that “despite the subjectivity involved in the analysis, the underlying theory of firearms identification is testable and the results have been verified.” *Id.* For example, the court in *United States v. Montiero*, stated that although an examiner’s opinion “is primarily subjective and based on the expertise of the examiner, the existence of requirements of peer review and documentation ensure sufficient testability and reproducibility to ensure that the results of the technique are reliable.” 407 F. Supp. 2d 351, 369 (D. Mass. 2006).

Since this Court’s *Wrensford* opinion and the publication of the PCAST Report, other courts have been critical of the field of firearms examination and have limited the testimony of firearms examiners. However, even courts that have been critical and have limited the testimony

have found the technique testable. *See United States v. Tibbs*, No. 2016-CF1-19431, 2019 WL 4359486, at \*7, \*23 (D.C. Super. Ct. Sep. 05, 2019) (finding that “virtually every court that has evaluated the admissibility of firearms and toolmark identification has found the AFTE method to be testable and that the method has been repeatedly tested,” but limiting the Government’s expert to testifying that “the recovered firearm cannot be excluded as the source of the cartridge casing found on the scene of the alleged shooting”). For example, in *United States v. Shipp*, the Eastern District of New York concluded that the “AFTE Theory can be and has been tested and this factor therefore weighs in favor of reliability,” even though the court precluded the Government’s expert from testifying “to any degree of certainty, that the recovered firearm [was] the source of the recovered bullet fragment or recovered shell casing.” 422 F. Supp. 3d 762, 776, 783 (E.D.N.Y. 2019).

The Court thus finds that there is no basis upon which to disturb its finding in *Wrensford* that the AFTE method is testable, and this factor weighs in favor of admissibility.

## **2. Peer Review**

“This *Daubert* factor focuses on whether the methodology employed by firearms and toolmark examiners is subject to peer review.” *Wrensford*, 2014 U.S. Dist. LEXIS 102446, at \*43. For this factor, Defendant does not argue that the AFTE methodology is not peer reviewed. Rather, Defendant argues that “the field lacks sufficient peer review and publication,” and that the studies presented by Dr. Hamby “contain clear statements of the authors’ biases, making clear the studies were conducted to validate the AFTE method for legal purposes rather than to test whether the method is valid.” (Dkt. No. 166 at 3; Dkt. No. 24 at 22). Defendant also argues that the *AFTE Journal* avoids scrutiny from scientists and academics who are not in the field of firearms examination because, generally, it is only available to members and others who pay a subscription

fee. (Dkt. No. 24 at 22).

In *Wrensford*, this Court found that the “AFTE theory is ‘subject to peer review through submission to and publication by the *AFTE Journal* of validation studies which test the theory.’” *Wrensford*, 2014 U.S. Dist. LEXIS 102446, at \* 43 (quoting *Otero*, 849 F. Supp. 2d at 433). The Court also noted that in addition to a formal submission process, the AFTE Journal had a “formal post-publication peer review process, allowing AFTE members and any other interested individuals to comment on previously published articles.” *Id.* (quoting *Otero*, 849 F. Supp. 2d at 433).

In court decisions since *Wrensford*, this factor has been found to weigh in favor of admissibility. In *Shipp*, the court found that prior court decisions had “with near uniformity, determined that the AFTE Theory [had] been subjected to peer review.” *Shipp*, 422 F. Supp. 3d at 776. The court in *Shipp*, however, expressed concerns about the *AFTE Journal*’s peer review process, in that reviewers were members of AFTE who had “a vested, career-based interest in publishing studies that validate their own field and methodologies.” *Id.* (quoting *Tibbs*, 2019 D.C. Super. LEXIS 4359486, at \*10). The *Shipp* court was also concerned that reviewers may not have “any specialized or even relevant training in research design and methodology.” *Id.* (quoting *Tibbs*, 2019 D.C. Super. LEXIS 4359486, at \*10).

Nonetheless, the *Shipp* court determined that the “AFTE theory [had] still been subjected to significant scrutiny”—which is the reason the Supreme Court found peer review important in *Daubert*—despite the *AFTE Journal*’s peer review process, and cited the PCAST report as proof. *See id.* (“*Daubert* found the existence of peer-reviewed literature important because ‘submission of the scrutiny of the scientific community . . . increases the likelihood that substantive flaws in the methodology will be detected.’”) (quoting *Daubert*, 509 U.S. at 593). The Court shares the

concerns expressed in *Shipp* and those of other courts, which have been critical of the *AFTE Journal*'s peer review process. *See, e.g., United States v. Adams*, 444 F. Supp. 3d 1248, 1265-66 (D. Or. 2020) (finding that the *AFTE Journal* was “a trade publication, meant only for industry insiders, not the scientific community” and “the purpose of publication in the *AFTE Journal* [was] not to review the methodology for flaws but to review studies for their adherence to the methodology”). Nonetheless, the Court finds that the AFTE methodology is peer reviewed, not solely based on the *AFTE Journal*, but based on the PCAST Report, 2009 NRC Study, and other publications, which have reviewed not merely the application of the AFTE methodology, but also the methodology itself. Indeed, in the *Daubert* hearing, Dr. Hamby testified that there were several peer-reviewed journals, one of which—the *Journal of Forensic Sciences*—published an article by Dr. Scurich. (Hr'g Tr. at 97).

Thus, the Court finds that the AFTE methodology has been subjected to sufficient peer review and publication, and this factor favors the admissibility of DeSouza's testimony.

### **3. General Acceptance**

The Court next examines whether firearms and toolmark analysis has achieved general acceptance in the “relevant scientific community.” *Daubert*, 509 U.S. at 594. Dr. Hamby testified that the field of firearms identification has been accepted by courts in the United States since at least the 1920s. *Id.* at 158. He also testified that, to his knowledge, it has never been rejected by any court. *Id.*

In *Wrensford*, this Court determined that firearms and toolmark identification was generally accepted. *Wrensford*, 2014 U.S. Dist. LEXIS 102446, at \*46; *see also Otero*, 849 F. Supp. 2d at 435 (“Courts have observed that the AFTE theory of firearms and toolmark identification is widely accepted in the forensic community and, specifically, in the community of



firearm and toolmark examiners.”). Although over the years courts have placed some limitations on the extent to which firearms examiners can testify to the existence of a match, such limitations have had no bearing on the general acceptance of the discipline. Thus, the Court finds no basis to disturb its determination in *Wrensford* and finds that this factor weighs in favor of admissibility.

#### **4. Standards Controlling the Technique’s Operation**

Defendant’s argument that the “sufficient agreement” standard is subjective attacks whether there are standards controlling the application of the AFTE methodology. (Dkt. No. 166 at 5). Defendant argues that the “method by which the examiners reach their conclusions is purely subjective and lacks a precisely defined process.” *Id.* at 1. Defendant states that the number of test fires used by laboratories varies; that there is “no set standard for how much of a bullet must be examined before making a determination”; and that there is “no set number of striated marks or impressed marks necessary to make an identification.” *Id.* at 8.

Defendant highlights that Dr. Hamby admitted that “aside from identifying class characteristics (which have objective, measurable, and definitive standards) firearms and toolmark comparisons are subjective.” *Id.* at 5. Defendant also cites to a 2008 study by the National Research Council, which found “that the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated.” *Id.* (citing National Research Council, *Ballistic Imaging* 3 (National Academies Press 2008) (“2008 NRC Study”)). Defendant argues that “no two fired ballistic items will match 100%,” and “markings made by a particular firearm on bullets and cartridge cases change over time.” *Id.* at 6. In further support, Defendant cites to Dr. Hamby’s testimony, where—for example—he explains that “you would never find two bullets that were absolutely identical. It’s impossible.” (Hr’g. Tr. at 236). Defendant also argues that the EBIS gun barrel was specifically designed because “[f]irearms examiners were not able to

distinguish which officers' firearms fired the fatal shot in officer involved shootings in Miami.” (Dkt. No. 166 at 6). The Fadul Study explains that there was an “inability to identify fired bullets to individual Glock pistols,” which resulted in the manufacturing of the EBIS barrel, and the EBIS barrels “present a significant advancement of polygonal rifled barrels.” Fadul, *supra* at 37.

The Court heard testimony from Dr. Hamby explaining that there are standards and procedures that firearms examiners use to perform their analyses. Examiners first ensure that the firearm is unloaded, safe and operable. (Hr’g Tr. at 64-65). The components are examined to determine “if they are the same caliber as the firearm,” and if they are not, they could not be fired from the same firearm. *Id.* at 65. The examiner will next do at least two test-fires, but some labs will do more. *Id.* at 67. Finally, the examiner looks “for replication of striae and/or impressed marks.” *Id.* at 82. Dr. Hamby testified that “you start seeing the individual marks . . . [a]nd you can say, yes, I see sufficient individual characteristics based on my training.” *Id.* at 83. Dr. Hamby also explained that firearms examination is subject to peer review, whereby a second examiner reviews the first examiner’s work, which is sometimes a blind review. *Id.* at 253.

In *Wrensford*, this Court discussed the protocols guiding firearms examiners, explaining that the firearms examiner followed protocols that comported “with industry standards for firearms examination,” which included peer review. *Wrensford*, 2014 U.S. Dist. LEXIS 102446, at \*49. There, as here, the Court found that there are standards controlling the analysis that firearms examiners undertake when analyzing bullets or cartridge casings. *See also Otero*, 849 F. Supp. 2d at 434 (detailing the steps outlined in a New Jersey State Police laboratory manual, which required examiners to “compare test-fired components against each other first under the comparison microscope to establish reproducibility of class and individual characteristics,” after which the examiner “may proceed to compare the discharged evidence . . . to a test fired shot known to have

originated from a particular firearm”).

Notwithstanding the Court’s finding that there are certain standards controlling the firearm and toolmark analysis, the Court agrees with the concerns Defendant raises here, regarding the circularity of the AFTE “sufficient agreement” standard. As the Court stated in *Wrensford*: “the AFTE ‘sufficient agreement’ criteria ‘does not provide any uniform numerical standard examiners can use to determine whether or not there is a match’ and therefore a conclusion that there is a match is ‘necessarily a subjective one.’” *Wrensford*, 2014 U.S. Dist. LEXIS 102446, at 53 (quoting *United States v. Taylor*, 663 F. Supp. 2d 1170, 1177 (D.N.M. 2009); *see also Shipp*, 422 F. Supp. 3d at 779 (stating that the “sufficient agreement” standard is circular, “subjective in nature,” and that different firearms examiners could have “different personal standards when two sets of toolmarks sufficiently agree”). Dr. Hamby stated that “[t]here’s subjectivity in most sciences,” including DNA analysis. (Hr’g. Tr. at 86). But the Court finds that “[t]here is a difference . . . between ‘some degree of subjective,’ as exists when a medical expert testifies as to whether a doctor met a certain accepted standard of care, and the near total subjectivity countenanced by the AFTE Theory, where there is no actual guidance for what comprises ‘sufficient agreement.’” *Shipp*, 422 F. Supp. 3d at 780; *see also Adams*, 444 F. Supp. 3d at 1263 (“Not only is the AFTE method not replicable for an outsider to the method, but it is not replicable between trained members of AFTE who are using the same means of testing.”).

Considering the Court’s findings regarding this factor and the level of subjectivity involved when the sufficient agreement standard is applied, the Court finds that the standards controlling firearms examinations militate in favor of admitting DeSouza’s testimony, but with limiting instructions.

### **5. Qualifications of the Expert Witness Testifying Based on the Methodology**

In *Wrensford*, the Court determined that “DeSouza’s knowledge, experience, and training qualify him to testify as an expert using the methodology described for firearms analysis.” *Wrensford*, 2014 U.S. Dist. LEXIS 102446, at \*55. The Court finds no reason to change its determination, and thus this factor—which was not challenged by Defendant—weighs in favor of admissibility.

### **6. The Known or Potential Rate of Error**

Defendant argues that “the field lacks a known error rate, as only one appropriately designed study—the [AMES I] Study—has attempted to measure it.” (Dkt. No. 24 at 23 (citing PCAST Report at 108-110)). Defendant argues that the AMES I Study “suggests that the field has grossly underestimated its rate of false positives.” *Id.* Further, “[w]ithout a known error rate, ‘an examiner’s statement that two samples are similar—or even indistinguishable—is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact.’” *Id.* (quoting PCAST, *supra* at 6). Defendant also argues that the studies relied upon by Dr. Hamby and the Government underestimate “the error rate because examiners can use the study design to deduce the correct answer.” (Dkt. No. 166 at 10). Further, Defendant explains that “with closed set-based studies . . . the researcher has no idea what comparisons were actually made as the examiner worked through the materials.” *Id.* Defendant argues that a false positive error rate cannot be calculated because “the researcher must know how many different source comparisons were made.” *Id.* at 11 (citation omitted). Defendant points to the PCAST Report, which found that “the closed-set design is problematic in principle and appears to underestimate the false positive error rate in practice.” *Id.* (citing PCAST Report, *supra* at 109).

Defendant also argues that in the open sample-based studies there were many more

“inconclusive” responses, which demonstrates that the “impact of being able to choose the closest ‘match’ in a closed set study is undeniable.” *Id.* at 13. Regarding, inconclusives Defendant also argues that there is “[n]o scientifically valid reason . . . to support the treatment of inconclusive responses as correct answers in studies.” *Id.* at 14. Rather, Defendant argues they should be treated as errors. *Id.*

The Court heard extensive testimony regarding the error rate for the field of firearms examination. Dr. Hamby’s and Dr. Scurich’s testimony on the subject focus on the following issues: (1) whether due to study design, the error rates in certain studies relied upon by Dr. Hamby accurately measure the error rates of firearms examiners when conducting casework; (2) how answers of “inconclusive” should be treated in calculating the error rates in studies; and (3) the actual error rate for the field of firearms examination.

Regarding the first issue, Dr. Hamby testified about the reported error rates in the following studies: his own study, the Keisler Study, the Mikko Study, the Fadul Study, the Thompson Study, the Brundage Study, the AMES I Study, and the AMES II Study. Dr. Scurich was critical of the study design for Dr. Hamby’s study, the Mikko Study, the Thompson Study, and the Brundage Study because they were all closed, set-to-set studies. In these studies, firearms examiners were given a set of unknown bullets and they were tasked with matching those bullets to known bullets, with each unknown bullet matching at least one known bullet. (Hr’g Tr. at 357). Dr. Scurich testified that based on the design of these studies, a false positive error rate could not be calculated because the number of different source comparisons was not known. *Id.* at 625. Dr. Scurich also testified that the study design enables examiners to match an unknown bullet to the closest known bullet because “even if a bullet doesn’t look exactly right but it’s very close, you could infer that it aligns with it.” *Id.* at 639. Dr. Scurich also criticized Dr. Hamby’s reliance on the Thompson

Study—on the grounds that it involved the identification of knives, not bullets or cartridge cases, *id.* at 558—and on the Fadul Study—on the grounds that it involved EBIS barrels, which are specially made to give each barrel its own unique signature and which are not involved in this case, *id.* at 535.

In his own study, Dr. Hamby reported seven inconclusives and no errors. *Id.* at 162. In the Mikko Study, the error rate reported was also zero. Mikko, *supra* at 291. In the Fadul Study, the reported error rate was 0.04 percent. Fadul, *supra* at 41. In the Thompson Study, the error rate reported was 0.776%. Thompson, *supra* at 369. In the Brundage Study, the error rate reported was zero. Brundage, *supra* at 440.

Regarding Defendant’s critique of the study design for the closed, set-to-set studies, the Court finds Defendant’s argument persuasive. As the PCAST Report states, the “‘closed-set’ design is simpler than the problem encountered in casework, because the correct answer is always present in the collection.” PCAST Report, *supra* at 109. In contrast, “in an open-set study (as in casework), there is no guarantee that the correct source is present—and thus no guarantee that the closest match is correct.” *Id.* at 108. When a firearms examiner is given a bullet or cartridge in casework, they do not know if it came from a given firearm, and so asking an examiner to match a bullet does not present an analogous situation to a closed-set study. *See Shipp*, 422 F. Supp. 3d at 778 (examining the PCAST Report and concluding that the closed-set design had advantages “none of which [were] present in fieldwork”).

Further, even though the Fadul Study was set-to-set and open, the Court shares Defendant’s concern that the study involves a specially made EBIS barrel, which is not present in this case. In addition, set-to-set study designs do not record the number of individual comparisons of known and unknown bullets/cartridge cases, which is what occurs in casework and would therefore

provide a more accurate calculation of an error rate. *See id.* (discussing the PCAST Report and explaining that for “black-box” studies “[t]he independence of each comparison means that an identification determination on one problem has no effect on the analysis of remaining problems” and such design “is the most similar to the situation examiners face in fieldwork”). Accordingly, the Court will consider the error rates in the sample-to-sample open studies—*i.e.*, the Keisler Study, AMES I and AMES II—to determine whether the error rate factor weighs for or against admissibility, because the design of these studies most closely approximates a firearm examiner’s casework.

Regarding the second issue, Dr. Scurich was critical of AMES I and AMES II because both studies counted “inconclusives” as correct answers, which decreased the error rate. (Hr’g. Tr. at 363, 377). Dr. Scurich argued that inconclusives should be treated as errors, “because [an inconclusive response] doesn’t match up with ground truth” and because permitting such responses effectively allows test participants “to pick and choose which questions they want to respond to.” *Id.* at 368, 378.

The Court agrees with Dr. Scurich that inconclusive responses are not properly classified as correct responses, because counting them as such would likely artificially *deflate* the error rate by counting two responses as correct and only one response as incorrect. However, counting inconclusive responses as errors is not a satisfactory solution because counting them as such would likely artificially *inflate* the error rate by counting two responses as incorrect and only one response as correct. Moreover, practical considerations also counsel against counting inconclusive responses as errors. As the PCAST Report correctly observes, if an examiner arrived at an inconclusive determination in casework, the evidence is unlikely to be used against a defendant in court. *See* PCAST Report, *supra* at 153 (calculating the error rate based on conclusive

examinations “because evidence used against a defendant will typically be based on *conclusive*, rather than inconclusive, examinations”). As such, counting inconclusive responses as errors has the effect of improperly equating inconclusive responses with the false positive or false negative errors that could adversely affect judicial proceedings by misleading the factfinder.

Accordingly, the Court will exclude inconclusive responses in the Keisler Study, and the AMES I and AMES II studies to determine the error rates, as done in the PCAST Report. *Id.* at 110. This results in an error rate of 0% for the Keisler Study, an error rate of 1.5 percent in AMES 1, and for AMES II, an error rate of 2 percent for bullets and 1.86 percent for cartridge cases. (*Id.*; Hr’g. Tr. at 377). Courts have found that error rates in this range caution “against the reliability of the AFTE Theory.” *Shipp*, 422 F. Supp. 3d at 778; *see also Adams*, 444 F. Supp. 3d at 1264 (considering toolmark error rate between 0.9 and 2.2 percent as problematic when put in the context of wrongful convictions, in that a 2.2 percent error rate “would mean that 1 in 46 convictions were wrong”). Other courts, however, have found these error rates to be acceptable. *See United States v. McCluskey*, 2013 U.S. Dist. LEXIS 203723, at \*26 (D.N.M. Feb. 7, 2013) (concluding that “insufficient data exist[ed] to calculate a definitive error rate,” but from available information “an error rate of 5% or less . . . [was] not excessively high”); *Romero-Lobato*, 379 F. Supp. 3d at 1120 (categorizing error rates for firearms examination ranging from 0.07 to 1.52 percent as “very low,” and finding the error rate to weigh in favor of admissibility).

The Court finds that an error rate ranging as high as two percent for the field is significant, but not so high as to weigh against the admissibility of DeSouza’s testimony entirely. Thus, this factor weighs in favor of admissibility but with limitations.



## **B. Admissibility of and Limitations on DeSouza's Testimony**

In sum, the Court finds that four factors weigh in favor of admitting DeSouza's testimony without limitation and two factors weigh in favor of admitting his testimony with limitations. Based on this, the Court will admit DeSouza's testimony but with limitations, due to the concerns discussed above, *i.e.*, the AFTE's "sufficient agreement" standard and the error rates from the relevant studies.

The Court's concerns, however, apply to the AFTE theory insofar as it is used to conclude that there is a match between an unknown bullet or cartridge case and one from a known firearm. Thus, DeSouza's testimony may include: (1) the theory of firearms and toolmark analysis; (2) the procedures he undertook to examine the cartridge cases, including inspecting and testing the firearm recovered, and his use of a comparison microscope to compare the cartridge cases to those test-fired from the recovered firearm; (3) whether the class characteristics matched; and (4) whether he found individual toolmarkings on the recovered cartridge cases to be consistent with those test-fired from the recovered weapon. DeSouza may also use comparison photographs to describe or show toolmarks from the recovered cartridge cases and those that were test fired. However, DeSouza may not testify as to a match between cartridge cases and the recovered firearm. *See Shipp*, 422 F. Supp. 3d at 783 (imposing same limitation); *Tibbs*, 2019 WL 4359486, at \*23 (imposing same limitation); *see also United States v. Davis*, No. 4:18-cr-00011, 2019 U.S. Dist. LEXIS 155037, at \*22-\*23 (W.D. Va. Sept. 11, 2019) (holding that toolmark examiners "may compare marks on various cartridge cases and identify marks on such cartridge cases they find to be similar and consistent with each other," but finding that "[c]oncerns over the reliability of this testimony expressed in the NRC and PCAST reports and those reflected in a recent chorus of federal decisions lead the court to impose certain restrictions," including that the examiners

could not testify “that the marks indicate a ‘match’” or “express any confidence level”).

“This limitation is in line with, albeit slightly more restrictive than, limitations that [some] other federal district courts have placed on toolmark analysis testimony.” *Shipp*, 422 F. Supp. 3d at 783. *Cf. United States v. White*, No. 17-CR-00611, 2018 WL 4565140, at \*3 (S.D.N.Y. Sept. 24, 2018) (precluding expert from testifying “to any specific degree of certainty as to his conclusion that there is a ballistics match”); *United States v. Glynn*, 578 F. Supp. 2d 56, 574-75 (S.D.N.Y. 2008) (limiting expert's testimony to stating that a match was “more likely than not”); *United States v. Simmons*, No. 16-CR-00130, 2018 WL 1882827, at \*8 (E.D. Va. Jan. 12, 2018) (limiting testimony to a “a reasonable degree of ballistic ... certainty”), *report and recommendation adopted*, No. 16-CR-00130, 2018 WL 658693 (E.D. Va. Feb. 1, 2018). The Court finds the limitation appropriate given the concerns regarding the AFTE’s “sufficient agreement” standard and the error rates in the field. Together with the testimony, as limited, the Court will allow DeSouza to use comparison photographs, as they would be helpful exhibits for the jury to view. *See Davis*, 2019 U.S. Dist. LEXIS 155037, at \*21 (“[T]here was no ‘reason why it would not be helpful to the jury for [the examiner] to testify with his photographs and matching up the marks he saw that were similar and pointing out the characteristics that were similar between the firearm cartridges fired at the scene and what was test fired, subject to cross-examination.” (quoting *United States v. Medley*, No. PWG 17-242 (D. Md. April 24, 2019), ECF No. 111, at 115)).

## V. CONCLUSION

For the reasons set forth above, the Court will grant in part and deny in part Defendant’s *Daubert* Motion. The Court will grant Defendant’s motion insofar as it will preclude DeSouza from testifying that the cartridge cases that were recovered matched, or came from, the firearm that was recovered, but the Court will otherwise deny Defendant’s motion. Thus, the Court will

permit DeSouza's testimony to include: (1) the theory of firearms and toolmark analysis; (2) the procedures he undertook to examine the cartridge cases, including inspecting and testing the firearm recovered, and his use of a comparison microscope to compare the cartridge cases to those test-fired from the recovered firearm; (3) whether the class characteristics matched; and (4) whether he found individual toolmarkings on the recovered cartridge cases to be consistent with those test-fired from the recovered weapon. Further, DeSouza may use comparison photographs to describe or show what he concludes were consistent toolmarks from the recovered cartridge cases and those that were test fired.

An appropriate Order accompanies this Memorandum Opinion.

Date: November 28, 2022

\_\_\_\_\_/s/\_\_\_\_\_  
WILMA A. LEWIS  
District Judge